**Note:** This document is a sample plan written before the NIH data management and sharing policy became effective and has not been reviewed by the NIH. Please direct questions or comments to Marla Hertz, mihertz@uab.edu.

**Data Management and Sharing Plan (NIH 2023)**

**ELEMENT 1: Data Type**

**1A. Types and amount of scientific data expected to be generated in the project:**

The proposed research will include data from approximately 60 postpartum women with overweight/obesity recruited from clinical facilities in Alabama. The final datasets will include quantitative outcomes including demographics, feasibility data, clinical data, and laboratory values, and other relevant measurements. Qualitative data will also be collected such as medical histories and interview responses from participants. The total amount of data generated is estimated to be in the GB range.

**1B. Scientific data that will be preserved and shared, and the rationale for doing so:**

Due to ethical considerations, the following data produced during the project will be preserved and shared:

1) Aggregate data on recruitment rates and reasons for refusal, retention rates, adherence rates, and treatment satisfaction (assessed both qualitative and quantitative) will be shared publicly.
2) De-identified IPD for the following measures will be restricted with a data use agreement: anthropometrics (weight, height, waist circumference), body composition (total and segmental body fat, lean mass, bone mineral density; visceral adipose tissue mass and volume), the following serum levels at each timepoint from the oral glucose tolerance test (OGTT): insulin, glucose, and c-peptide; dietary data (24-hour energy intake, macronutrient and micronutrient intake), visual analogue scale scores for appetite, 8-item Patient-Reported Outcomes Measurement Information System® (PROMIS) Fatigue Short Form (SF) 8a (v1.0) survey responses and total score, Pittsburg Sleep Quality Index survey responses and total scores, International Physical Activity Questionnaire-Long Form (IPAQ-LF) survey responses and total scores, Edinburgh Postnatal Depression Scale responses and total score, breastfeeding duration, breastfeeding intensity (exclusive breastfeeding, mixed feeding, exclusive formula feeding), and reasons for discontinuing breastfeeding, breast milk energy density, breast milk total fat, infant recumbent length, infant weight, infant z-scores calculated according to World Health Organization Standards, demographics (age, sex, annual household income, employment status, educational attainment, race and ethnicity, health insurance coverage, number of adults in the household, number of children in the household), and medical history (parity, gravidity, pre-pregnancy weight, gestational weight gain, gestational age at delivery, medical conditions). See Table 1 for overview of the sharing plan.

Table 1: Data sharing overview

| Measure | Level shared* | Data standard |
|---|---|---|
| Process data | | |
| Recruitment | aggregate | Best practices |
| Retention | aggregate | Best practices |
| Adherence | aggregate | Best practices |
| Treatment Satisfaction | aggregate | Best practices |
| Clinical outcome data | | |
| Weight | IPD | USCDI |
| Body Composition | IPD | USCDI |
| 2-hour OGTT | IPD | USCDI |
| Patient-reported outcome data | | |
| Energy intake | IPD | Best practices |
| Appetite | IPD | Best practices |
| Fatigue | IPD | Best practices |
| Other measures | | |
| Sleep | IPD | Best practices |
| Physical Activity | IPD | Best practices |
| Depression | IPD | Best practices |
| Breastfeeding | IPD | Best practices |
| Breast milk composition | IPD | Best practices |
| Infant anthropometrics | IPD | USCDI |
| Demographic Data | IPD | USCDI |
| Medical History | IPD | USCDI |

*Aggregate, shared publicly; IPD, controlled access

**1C. Metadata, other relevant data, and associated documentation:**

To facilitate interpretation of the data, metadata including a data dictionary, statistical analysis plan, analytic code, and final protocol will be shared and linked to relevant datasets.

**ELEMENT 2: Related Tools, Software and/or Code**
De-identified clinical and patient-reported data will be made available in a .csv file and will not require the use of specialized tools to be accessed or manipulated. Some file types will require the use of specialized software. Quantitative data analysis will generate SAS files (.sas7bdat) which requires SAS Analytics Software (SAS Institute) to access and manipulate. Data from audio recorded during qualitative interviews will be stripped of any identifiers and written transcripts will be coded and analyzed using NVivo. Access to NVivo files (.nvp) will require NVivo software (QSR International).

**ELEMENT 3: Standards**
To facilitate interpretation and interoperability, all data and materials will be structured and described using the following standards (see Table 1). The study will maximize use of common data elements and standard survey instruments. We will follow U.S. Core Data for Interoperability (USCDI) for clinical data and laboratory values. Proper data input will be assured by using REDCap software. Formal standards for some have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices.

**ELEMENT 4: Data Preservation, Access, and Associated Timelines**
**4A. Repository where scientific data and metadata will be archived:**
All aggregate datasets that can be shared publicly will be deposited in the Open Science Framework (OSF) generalist repository. Individual participant data (IPD) will be shared via controlled access in the Vivli repository.

**4B. How scientific data will be findable and identifiable:**
Both Vivli and OSF provide metadata, persistent unique identifiers, and guarantee long-term access of ≥ ten years.

**4C. When and how long the scientific data will be made available:**
Data will be made available at the time of associated publication or by the end of the clinical trial.

**ELEMENT 5: Access, Distribution, or Reuse Considerations**
**5A. Factors affecting subsequent access, distribution, or reuse of scientific data:**
Due to the sensitive nature of the study, the type of sharing depends on the level of data aggregation. Individual survey responses will not be shared to protect the privacy of research subjects. Medical history data, such as the date the participant most recently gave birth, and any of the 18 HIPAA identifiers (e.g., names, geographic subdivisions smaller than a state, all elements of dates directly related to an individual, etc.) will not be shared. Additionally, audio recordings of qualitative interviews will not be shared, as they could be used to identify participants in conjunction with quantitative data. Deidentified IPD data for the four data types will be shared by controlled access as outlined in 5B.

**5B. Whether access to scientific data will be controlled:**
Deidentified IPD data will be deposited in the Vivli data repository, which restricts access to the data to qualified investigators with an appropriate research question who sign a data use agreement (DUA). The Vivli DUA limits subsequent use to the terms of the approved request and requires that users maintain data security, and refrain from any attempts to reidentify research participants or engage in any unauthorized uses of the data. To request access to the data, the user must submit a valid scientific question, include a statistical analysis plan, and complete all required fields on the Vivli data request form. Vivli will review the data request for completeness. Anyone who has submitted an approved data request and signed a data use agreement on Vivli will be given access to the data without cost, for a set period.

**5C. Protections for privacy, rights, and confidentiality of human research participants:**
Data will be de-identified by the NLM Scrubber and expert determination methods before it is shared to protect participant privacy. The IRB protocol and informed consent documents will include language describing the data management and sharing plan, explaining the motivation for sharing, and ensuring that personal identifying information will be removed prior to sharing.

**ELEMENT 6: Oversight of Data Management and Sharing**
The PI and a project manager will be responsible for educating the research team on the data management plan and ensuring team member compliance during the project. The PI will be responsible for updating and revising the Data Management and Sharing Plan when necessary and reporting compliance in annual RPPR.